Tec Empresarial
P-ISSN: 1659-2395; E-ISSN: 1659-3359

HARMONIZING EMOTION: A MULTIMODAL APPROACH TO ANALYZING HUMAN AFFECT IN MUSIC RECOMMENDATION SYSTEMS

Mrs. Ketaki A. Bhosale[1] , Dr. Sangram T. Patil[2]
Assistant Professor[1], Associate Professor[2]
Department of Computer Science & Engineering[1,2]
D Y Patil College of Engineering Technology Kolhapur, Maharashtra, India[1], DYP-ATU, Talsande, Maharashtra, India[2]
ketakibhosale28@gmail.com[1], sangrampatil@dyp-atu.org[2]

Abstract:

Music is a worldwide language that everyone throughout the world enjoys. Zatorre and Peretz (2001) state that, musical undertakings with their unique essence appear to have been a part of every recorded society on Earth, dating back at least 250,000 years [1]. As the digital age advances, Customized music suggestion systems are now deeply ingrained in our everyday routines, providing us with a curate's selection of songs that match our preferencesMudit Kumar Tyagi et.al [2]suggested a method for extracting user preferences based on their music listening history.Incorporating demographic information such as age and gender provides a more nuanced understanding of a listener's identity. People of different ages and genders may have unique musical preferences, and these attributes can act as significant filters in the recommendation process. For example, a teenager's taste in music is likely to differ from that of a middle-aged adult. Similarly, gender can play a role in shaping musical choices. Integrating age and gender detection into music recommendation systems ensures that the music offered is not only personally relevant but also age-appropriate and respectful of gender sensitivities. This research proposes a multimodal approach, combining demographic human features and emotional signals, to refine and personalize music selection through advanced machine learning techniques.

Keywords: Music recommendation, Multimodal approach, demographic features

Introduction

The impact of music on human behavior is multifaceted. Studies indicate music significantly influences emotions, spanning joy, excitement, sadness, and nostalgia. Music therapy is employed clinically to address anxiety, depression, and stress, enhancing emotional well-being. Music shapes and reflects cultural identity, influencing social norms and values. Physiologically, music impacts heart rate, blood pressure, and cortical levels.

Personalized traditional music recommendation systems aim to provide users with music suggestions tailored to their individual preferences. Various approaches and algorithms such as

Collaborative Filtering,Content-Based Filtering,Hybrid Systems,Matrix Factorization,Deep Learning Models,Knowledge-Based Systems,Context-Aware Recommendation,and Implicit Feedback Models are used individually or in combination depending upon available data, system goals, and the desired level of personalization.

From the various mentioned approaches there are three main approaches to customize music recommendations: collaborative filtering (CF) [3], content-based (CB) [4], and hybrid [5]. Based on the songs that users have listened to in the past, CB suggestions present similar songs to them. CF recommendations make music recommendations to users based on an analysis of the listening preferences of people with similar tastes. The hybrid method combines the insights from both the CF and CB methodologies to provide personalized music recommendations. Following table compares the three approaches in music recommendation system

Music Recommender System	**Data Working Source**	**Technology Used**	**Website**
Content Based Recommendation	Uses the user's historical data and takes into account the audio's inherent characteristics.	Gaussian Mixture Models (GMM) & Word Frequency Mining (WFM)	Shrimps Music
Collaborative Filtering Recommendation	Consider the users rating for a particular music.	Association Rule, KNN, Clustering, DecisionTree, Regression, CNN	Last FM music station
Hybrid Approach Recommendation	Combines the approaches of different music recommendation systems	Combination of content and collaboration Techniques	7HCCMR

Table 1: Summary of different approaches for Music Recommendation System

Demographic features, such as age, gender, and emotion can be valuable in understanding and enhancing personalized music recommendations.Considerable research has utilized deep learning techniques like Convolutional Neural Networks (CNN) and Artificial Neural Networks (ANN) for age, gender estimation, and emotion detection. Within CNN, Feature extraction identifies age, gender, and emotion-related features. Additionally, Feature classification in CNN categorizes facial images accurately into age groups, genders, and emotions such as happiness, sadness, anger, and neutrality.

The music recommendation problem can be divided into two sub problems first is Forecasting i.e predicting the likely music for a user and second is recommending or suggesting the list of probable music, the user loves to listen.

Forecast: Let I = $\{i_1, i_2,, i_n\}$ be the set of all possible items that can be recommended (a goal music collection), and let U = $\{u_1, u_2,..., u_m\}$ be the set of all users. Every user interface demonstrated interest in a certain set of goods. $Iu_i \subseteq I$.

Suggestion: Calculate the function $P_{ua,\ ij,}$an anticipated preference which denotes that itemij$\notin I_{ua}$ for the active user u_a

Literature Review:

Research in various areas is made to detect age, gender and emotions of the people. While some of the researchers took audio, others used image capture to extract features before conducting analysis. The following table depicts the summary for publication papers related to age, gender & emotion detection systems.

Paper No	Objectives	Method Used	Findings
[6] (2023), [7] (2022), [8](2020), [9](2020).	The main focus of the article is on using facial photos to assess age, gender, and emotions in real time.	The HOG-Viola-Jones algorithm demonstrates high accuracy in age, gender, and emotion recognition. •Recognition tasks employed CNN along with algorithms such as AdaBoost, PCA, HOG, LBP, HAAR, FPLBP, and LDA.	• A complete survey of techniques for age, gender, and emotion classification was reviewed • The Viola-Jones algorithm serves for object detection and face detection purposes. Local Binary Pattern (LBP) finds application in texture classification and real-time image analysis.

[10](2022) [11](2020) [12](2016)	• The paper compares nine conventional learning methods for mood detection. Mood classifiers' efficacy is evaluated using Twitter data pertaining to COVID-19. System recommends songs based on user's mood and preferences	• The utilized methods encompass Complement Naive Bayes, Random Forest, Decision Tree, and classifiers. • Gaussian classifiers ,Multi-class and rule-based, Bayesian were employed. • Binary classifiers were adapted for multi-class categorizations.	• Complement Naive Bayes outperforms Random Forest and Decision Tree in detecting mood variations. • Simple classifiers can be used for studying mood patterns in individuals. • Deep learning algorithms can be studied for classifying text based on moods.

[13](2020) [14](2010) [15](2018)	• The aim of the paper is to detect emotions in real-time using webcam images. • Features are extracted from facial landmarks for emotion detection.	ReLU(Rectified Linear Unit), CNN(Convolutional Neural Network), Max-Pooling Circular Local Binary Pattern, KNN(K-Nearest Neighbors) Logistic Regression Image Recognition Feature Extraction	• Proposed model predicts sentiment based on video information. • Resulting output can be used to address mental disorders and stress.
[2](2014)	Music Information Retrieval (MIR) was designed using two case studies	Two case studies, Emotify and Hooked, were established for gathering data in the field of Music Information Retrieval (MIR). Emotify specializes in emotional annotation of music. Hooked explores musical catchiness.	● Collecting data through online multiplayer games for music research. ● Developing games to annotate music emotionally and investigating musical catchiness.

[17](2015)	The voice-based speaker processing system is investigating speaker attributes such as age and emotions (including stress and mood), which may vary depending on gender.	Principal component analysis (PCA), Meel frequency cepstral coefficients (MFCCs),Gaussian mixture model (GMM)	The system discerns the age and emotions of speakers, considering gender differences. The proposed system aims to enhance human-computer interaction.
[18](2020)	To achieve the highest accuracy in predicting emotions among individuals experiencing depression.	Meel frequency cepstral coefficients (MFCC), Multi-layer Perceptron classifier (MLPC)	The framework preprocesses the audio data and identifies emotions using the MLP classifier.

[19](2014)	Introduces a system capable of discerning an individual's emotional state from recorded audio signals.	Support Vector Machine (SVM) classifiers ,Pitch Frequency Estimation method	The system is composed of two subsystems: 1) emotion recognition (ER) 2) gender recognition (GR) The experimental findings underscore that integrating the Gender Recognition (GR) subsystem enhances the overall accuracy of emotion recognition from 77.4% to 81.5%.

Table 2: Summary for age, gender and emotion recognition

Proposed System

The proposed system can identify emotions more accurately by combining information from multiple modalities, including text analysis, speech tonality, and facial expressions. When combined, the distinct insights from each modality can provide a more thorough picture of the user's emotional state.The accuracy, resilience, and user experience of the system can be greatly improved by using a multimodal approach to emotion recognition and music selection. This will result in interactions that are more engaging, natural, and sympathetic.

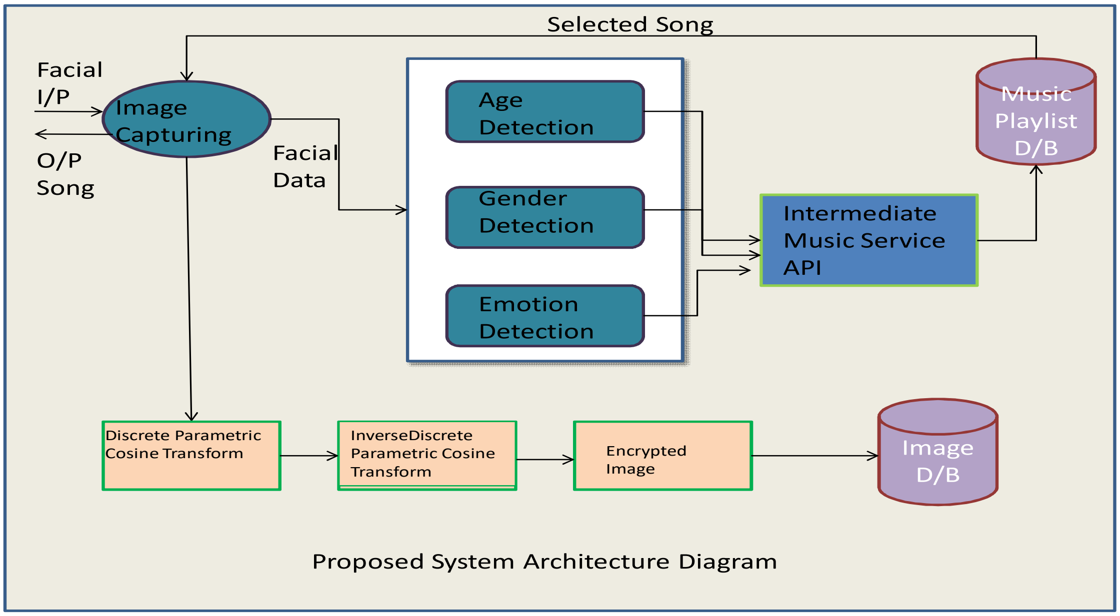

Fig 1: Proposed System Architecture Diagram

The Proposed system can be explained using following terms:

Multimodal Approach:A multimodal approach involves combining information from multiple modes or modalities to enhance understanding, representation, or interaction in a system. The study encompasses the collection of diverse user data such as age ,gender along with emotional signals captured using multiple modalities, such as facial expressions, physiological responses, and user-provided emotional labels this data can be audio ,video or text.

Dataset creation:

The image, gender and emotion dataset will be collected from IMDB-WIKI Dataset , LFW (Labeled Faces in the Wild) Dataset ,AffectNet Dataset. After working on these three datasets some real time images of people with different ages , gender and various facial expressions will be collected.

Age Detection:

Machine learning and deep learning models will be utilized for age detection. Common deep learning architectures such as Convolution Neural Networks (CNNs) or recurrent networks can be utilized for detecting age from both image and voice data.A deep learning architecture called a convolution neural network (CNN) is made to learn straight from data. It is a kind of artificial neural network that is widely used for object and picture categorization and recognition. Deep Learning is able to recognize things in photos by using CNNs. These networks are essential for many applications, including speech recognition in natural language processing, video analysis, obstacle recognition in autonomous cars, image processing, and computer vision tasks like segmentation and localization.

In a CNN, the input layer receives image pixels arranged in arrays. Multiple hidden layers within CNNs engage in feature extraction from the image through various operations such as convolution, pooling, rectified linear units, and fully connected layers. The convolution layer initiates the feature extraction process from the input image. Finally, the fully connected layer categorizes and identifies the object, producing the output layer.

Gender and Emotion Detection:

Machine learning models, such as deep learning models like CNNs or Support Vector Machines (SVM) for gender detection from audio and visual data is widely used. For Emotion Detection CNNs for facial expression analysis.

Music recommendation algorithms:

After identification of the age, gender along with emotions such as happiness, surprise, anger, neutrality, and sadness The system provides a curates playlist of music that matches the detected mood and the other parameters. Algorithms like collaborative filtering, content-based filtering, and hybrid methods, take into account the user's age, gender, and emotional state to generate personalized music recommendations.

Providing user privacy & Feedback Mechanism:

Implement strict privacy protection mechanisms to safeguard user data of captured image used to detect gender, age, and emotion as this data can be sensitive. The captured image can be encrypted Discrete Parametric Cosine Transform (DPCT) algorithm. The 2D DPCT, a sophisticated cosine transform, necessitates 12 parameters, posing challenges in real-world applications. Nonetheless, these parameters enhance the potency of the 2D DPCT, furnishing it with robust characteristics.

Summary :

This review study delves into the nascent domain of multimodal techniques in music recommendation systems, emphasizing three crucial demographic aspects: age, gender, and emotion. It compiles recent findings and approaches that use several modalities, including user listening history, lyrics, audio content analysis, and contextual information from social media. Accurately recognizing gender, emotional state, and age group through these multimodal methodologies presents both potential and challenges; advances in machine learning algorithms and feature extraction techniques are highlighted. It also looks at the effects of using this kind of demographic data in music recommendation algorithms, such as better user experiences and customized playlists. All things considered, the research highlights how multimodal methods can be used to customize music recommendations based on complex demographic preferences, opening the door to more advanced and user-focused music recommendation systems.

References:

[1] I. Eibl-Eibesfeldt, “The Biological Foundation of Aesthetics,” Beauty and the Brain, pp.29–68, 1988, doi: 10.1007/978-3-0348-6350-63

[2]Mudit Kumar Tyagi , Muhammad Ali , Garvit Kaim , Tripti Lamba , Gunjan Chugh ,”Music Recommendation System Using Multiple Machine Learning Models” International Conference on Innovative Computing & Communication (ICICC) 2023

[3] Content-Based Recommendation Systems Michael J. Pazzani& Daniel Billsus Part of the book series: Lecture Notes in Computer Science ((LNISA,volume 4321))

[4] J. Ben Schafer, D. Frankowski, J. Herlocker, and S. Sen, “LNCS 4321 - Collaborative Filtering Recommender Systems,” no. January 2007, 2014.

[5] Hybrid Web Recommender Systems Robin Burke Part of the book series: Lecture Notes in Computer Science ((LNISA,volume 4321))

[6] Smart Facial Emotion Recognition With Gender and Age Factor Estimation Surya Teja Chavali a, *, Charan Tej Kandavallib , Sugash T M c , Subramani R

[7]An Approach Based on Deep Learning for Recognizing Emotion, Gender and Age

Nippon Datta Nippon1 Juel Sikder Juel1

[8] Age Prediction using Image Dataset using Machine Learning Vijay Kumar

[9] REGA: Real-Time Emotion, Gender, Age Detection Using CNN—A Review Dibya JYOTI Sharma Kaziranga University , Sachin Dutta

[10]Mood detection and prediction using conventional machine learning techniques on COVID19 data Subhayan Bhattacharya1 · Abhay Agarwala1 · Sarbani Roy

[11]Emotional Detection and Music Recommendation System based on User Facial Expression

S Metilda Florence1 and M Uma2

[12]Emotion Based Music Player – XBeats Aditya Gupte, Arjun Naganarayanan, Manish Krishnan

[13]Human Emotion Detection using Machine Learning Techniques PunidhaAngusamy

[14]Biased emotional recognition in depression: Perception of emotions in music by depressed patients MarkoPunkanen, Tuomas Eerola, and Jaakko Erkkilä.

[15] An Emotion-Aware Personalized Music Recommendation System Using a Convolutional

Neural Networks Approach Ashu Abdul Jenhui Chen Hua-Yuan Liao

[16] Designing Games with a Purpose for Data Collection in Music Research. Emotify and Hooked: Two Case Studies AnnaAljanaki, Dimitrios Bountouridis, John Ashley Burgoyne, Jan Van Balen, Frans Wiering, Henkjan Honing, and Remco Veltkamp.

[17] Automatic Speaker Age Estimation and Gender Dependent Emotion Recognition Shivaji J. Chaudhari Ramesh M. Kagalkar

[18]Emotion and Stress Recognition through Speech Using Machine Learning Models Druva Manasa and C. Kiran Ma

[19] Gender-Driven Emotion Recognition Through Speech Signals for Ambient Intelligence Applications IGOR BISIO, ALESSANDRO DELFINO, FABIO LAVAGETTO, MARIO MARCHESE, AND ANDREA SCIARRONE